a characteristic comparison routine on the server, identifying a characteristic of the file content based on the appearance of the file content ID in the appearance database and transmitting the characteristic to the client agents.

32. The content classification system of claim 31 wherein said ID generator comprises a hashing algorithm.

33. The content classification system of claim 32 wherein said hashing algorithm is the MD5 hashing algorithm.

34. The content classification system of claim 31 wherein said ID appearance database tracks the frequency of appearance of a digital ID.

35. The content classification system of claim 31 wherein said plurality of agents are coupled to said database via a combination of public and private networks.

36. The content classification system of claim 35 wherein said database is coupled to an intermediate server which is coupled to said plurality of agents.

37. The content classification system of claim 36 wherein said intermediate server is a web server.

38. The content classification system of claim [39] 31 wherein said characteristic comprises junk e-mail and said characteristic is defined by a frequency of appearance of a file content ID.

39. A method for identifying characteristics of data files, comprising:

receiving, on a processing system, file content identifiers for data files from a plurality of file content identifier generator agents, each agent provided on a source system and creating file content IDs using a mathematical algorithm, via a network;

determining, on the processing system, whether each received content identifier matches a characteristic of other identifiers; and

outputting, to at least one of the source systems responsive to a request from said source system, an indication of the characteristic of the data file based on said step of determining.

40. The method of claim 39 wherein said file content identifier generates an identifier by hashing at least a portion of the data file.

41. The method of claim 40 wherein said hashing comprises using the MD5 hash.

42. The method of claim 40 wherein said step of generating comprises hashing multiple portions of the data file.

43. The method of claim 39 wherein each said data file is an email message and said step of determining comprises determining whether said email is SPAM.

44. The method of claim 39 wherein said step of determining identifies said e-mail as SPAM by tracking the rate per unit time a digital ID is generated.

45. The method of claim 44 wherein said method further includes the step of instructing said plurality of source systems to perform an action with the email based on said determining step.

46. A method of filtering an email message, comprising:

receiving, on a second computer, a digital content identifier created using a mathematical algorithm unique to the message content from at least two of a plurality of first computers having digital content ID generator agents;

comparing, on the second computer, the digital content identifier to a characteristic database of digital content identifiers received from said plurality of first computers to determine whether the message has a characteristic; and

responding to a query from at least one of said plurality of computers to identify the existence or absence of said characteristic of the message based on said comparing.

47. The method of claim 46 wherein said second computer is coupled to said plurality of first computers by a combination of public and private networks.

48. The method of claim 47 wherein said step of receiving includes receiving identifiers from said plurality of first systems via an intervening Web server.

49. The method of claim 48 wherein said plurality of systems are coupled by the Internet.

50. The method of claim 46 wherein said step of comparing comprises determining the frequency of a particular ID occurring in a time period, classifying said ID as having a characteristic, and comparing digital content identifiers to said classified IDs.

Please cancel claims 51 – 54.

55. A file content classification system for a first computer and a second computer coupled by a network, comprising:

a client agent file content identifier generator on the first computer, the file **content identifier comprising a computed value of at least two non-contiguous** sections of data in a file; and

a server comparison agent and data-structure on the second computer receiving identifiers from the client agent and providing replies to the client agent;

wherein the client agent processes the file based on replies from the server comparison agent.

56. A method for providing a service on the Internet, comprising:

collecting data on a processing system from a plurality of systems having a client agent generating digital content identifiers created using a mathematical algorithm for each of a plurality of files on the Internet to a server having a database;

characterizing the files on the server system based on said digital content identifiers received relative to other digital content identifiers collected in the database; and

transmitting a [content] substance identifier from the server to the client agent indicating the presence or absence of a characteristic in the file.

57. The method of claim 56 wherein said step of collecting comprises collecting a digital identifier for a data file.

58. The method of claim 57 wherein said file content is an e-mail.

59. The method of claim 57 wherein said step of characterizing comprises:
tracking the frequency of the collection of a particular identifier;
characterizing the data file based on said frequency;
storing the characterization; and
comparing collected identifiers to the known characterization.